A spatial analysis of Italian unemployment differences

Cracolici, M.F.; Cuffaro, M.; Nijkamp, P.

published in
Statistical Methods and Applications
2009

DOI (link to publisher)
10.1007/s10260-007-0087-z

document version
Publisher's PDF, also known as Version of record

Stat Methods Appl (2009) 18:275–291
DOI 10.1007/s10260-007-0087-z

OpenAccess

ORIGINAL ARTICLE

A spatial analysis on Italian unemployment differences

Maria Francesca Cracolici · Miranda Cuffaro · Peter Nijkamp

Accepted: 27 November 2007 / Published online: 3 January 2008

Abstract Using spatial econometric models, this paper focuses attention on the spatial structure of provincial unemployment disparities of Italian provinces for the year 2003. On the basis of findings from the economic literature and of the available socio-economic data, various model specifications including supply- and demand-side variables are tested. Further we use ESDA analysis as equivalent to integration analysis on time series; therefore it is applied on each variable, dependent and independent, involved in the statistical model. The suggestions of ESDA lead us to the most adequate statistical model, which estimates indicate that there is a significant degree of neighbouring effect (i.e. positive spatial correlation) among labour markets at the provincial level in Italy; this effect is present notwithstanding we controlled for local characteristics. The unemployment shows a polarized spatial pattern that is strongly connected to labour demand and to a much lesser extent to the share of young population and economic structural composition.

Keywords Regional unemployment · ESDA · Spatial models

1 Introduction

Geographic unemployment rates are often regarded as signposts for the socio-economic performance of regions. And, consequently, the analysis of regional unemployment differences has attracted increasing interest in the economic literature.

M. F. Cracolici (✉) · P. Nijkamp
Department of Spatial Economics, Free University of Amsterdam,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
e-mail: mcracolici@feweb.vu.nl

M. Cuffaro
Department of National Accounting and Analysis of Social Processes, University of Palermo,
Viale delle Scienze 28, Palermo 90128, Italy

There is an abundance of empirical literature that tries to explain the differences between geographical areas in terms of unemployment rates (see, e.g., Fischer and Nijkamp 1987; Decressin and Fatás 1995; Jimeno and Bentolila 1998; López-Bazo et al. 2002, 2005).

The principal aims of the empirical literature on regional unemployment are to examine the persistence of unemployment differentials and to define a model that investigates its determinants. The analyses contain several complementary and sometimes contrasting frameworks and have brought to light some interesting stylized facts, notably: (a) regional labour markets in Europe and the USA differ significantly; (b) regional differences in unemployment in European regions are more persistent than in the USA; (c) the persistence of unemployment differences in European regions is mainly due to poor flexibility of wages and low mobility of workers. In particular, in Italy, as in several other European areas, the persistence of unemployment is due to both structural problems and the inability of Italian regions to absorb specific shocks (on the demand or the supply side) (for details, see Dohse et al. 2002).

Jimeno and Bentolila (1998), following the findings of the most noteworthy literature on the US and EU labour market, classified the areas of unemployment into three groups, on the basis of the degree of persistence of the aggregate and regional relative unemployment: (1) low persistence of aggregate and regional relative unemployment (this is the case for the USA); (2) high and low persistence of, respectively, aggregate and regional relative unemployment (this is the case for most regions of the EU); (3) high persistence of aggregate and regional relative unemployment (this is the case of some European countries like Italy or Spain). The main reason for this strong stability and persistence of high regional unemployment may be connected to the low flexibility of the labour market, i.e. wage rigidity and employment inertia, and to labour force dynamics.

In particular, with regard to Britain and Italy, Eichengreen (1992) argued that the responsiveness of migration to a regional labour market disequilibrium is greater in the US than in either of the other two countries. And hence, it is reasonable to hypothesize that the adjustment process towards the equilibrium-state is due to other mechanisms (e.g. wage adjustment, labour-leisure choice, interregional capital mobility, etc.) which balance the limited mobility of workers.

The previous literature stressed the different behaviour of the Italian labour market with respect to both other European countries and the USA. This different performance of the Italian labour market leads us to investigate it at a more detailed territorial level. A stylized fact of the Italian labour market is the North–South dichotomy (see, e.g., Faini et al. 1997; Prasad and Utili 1998; Brunello et al. 2001). However, this dichotomy actually hides a patchwork of local facts that could be better explained by a provincial analysis (see Gambarotto and Maggioni 2002).

On the basis of the above considerations we propose to analyse the geographical distribution of Italian unemployment by using a single equation model.

Most of the theoretical and empirical literature explains and interprets unemployment differences by starting from the hypothesis of a stable equilibrium of spatial labour markets (Hall 1970; Marston 1985).

More specifically, when the effect on local or regional unemployment—caused by short run shocks—is dissipated, the persistence of differences in unemployment rates can be interpreted in terms of equilibrium (or disequilibrium in nature) (Marston 1985). According to this interpretation, the unemployment in each area is a function of the amenities and the endowments of the land. Workers migrate to areas where new jobs are created until there is no further incentive to move. In other words, the spatial distribution of unemployment under an equilibrium interpretation is characterized by constant utility across areas: high unemployment in the ith area is compensated for by some other positive factors (e.g. local amenities, climatic conditions, quality of life, local housing prices, etc.) which are a disincentive to migration. Similar considerations can be put forward with regard to firm migration.

In contrast to the previous interpretation, local unemployment differentials can also be explained in terms of disequilibrium. The disequilibrium view assumes that in the long run the unemployment rate will level off across areas. The adjustment process may be faster or slower and, depending on its speed, differences in unemployment across areas could persist for a long time. The speed of the adjustment may depend on different factors connected to both labour demand and supply.

Based on this previous interpretation of labour market and recent cross-sectional analyses, our paper explores the differences of unemployment across provinces in terms of some specific characteristics of local areas (e.g. housing, sectorial composition, etc.); local population and labour demand (see, e.g., Molho 1995; Patridge and Rickman 1997; Niebuhr 2003; López-Bazo et al. 2002, 2005). In the model proposed, the unemployment is explained by a proxy for labour demand and by some control variables; moreover, to take into account the neighbouring effects of labour markets (see Overman and Puga 2002) spatial econometric tools have been used.

The applied analyses are mainly based on time series, using standard statistical methods, both parametric and non-parametric (see Decressin and Fatás 1995; Jimeno and Bentolila 1998; Martin 1997; López-Bazo et al. 2002). There are only a very few analyses using spatial data and spatial parametric tools (see Molho 1995; Aragon et al. 2003; Niebuhr 2003; López-Bazo et al. 2002).

We will use spatial econometric methods based on spatial autocorrelation techniques to explore the geographical distribution of unemployment for the 103 Italian provinces for the year 2003. The aim of the paper is to construct a model that is capable of both incorporating the main findings of this literature and the statistical characteristics of data, viz. the spatial structure. This is detected by the ESDA analysis used to collect indications on which variable could be inserted in the model. Further, we are interested to test whether the time persistency of unemployment, as empirically supported by much Italian research (see Contini and Trivellato 2005), corresponds to a persistence of unemployment in space (i.e. neighbouring provinces tend to have similar unemployment rates). We will explore whether the unemployment in the provinces in Italy has a uniform spatial structure with very few *pockets of nonstationarity* (see Anselin 1995): in other words, if the provinces with high unemployment rates are contiguous (or share a border) with those provinces with low unemployment rates. As far as we know, this is the only empirical spatial analysis on the Italian labour market.

The paper is structured as follows. Section 2 presents the statistical models and the data used in our empirical application. In Sect. 3, the empirical findings are presented and interpreted. And, finally, some concluding remarks are made in Sect. 4.

2 Models and data on unemployment in Italian provinces

The aim of the present analysis is to investigate whether there is a spatial relationship among provincial unemployment disparities in Italy. Using observations on a provincial level, the present analysis emphasises the territorial aspect of labour markets and explores in more detail stylized facts on the dual unemployment structure of the Italian market (the North–South dichotomy), in order to highlight the differences between the complex and varied structure of local labour markets.

According to the literature, regional differences of unemployment and its spatial patterns may be explained by differences at the local level due to structural and non-coincidental factors. Usually, the variables used involve the following aspects: natural change, participation, migration, commuting, wages, unionization, employment, gross regional product, market potential, size and density, industry-mix explanation, economic and social barriers, and educational attainment of the population (for details, see Elhorst 2003). Among these aspects related to labour market there are some particularly recurrent in empirical analysis of unemployment differences. Specifically, it is quite common to relate the unemployment rate to the employment and economic structure as proxies of labour demand; population structure of local area as proxy of labour supply (i.e. age population, gender composition, etc.). Other explanatory variables could be introduced to take into account the amenities (e.g. housing, density) and the availability of workers to move (e.g. commuting).

Because of lack of data, the present analysis considers only some proxies of the previous features which characterize the local labour market. Using spatial econometric models we investigate the significance of spatial interaction of unemployment disparities in the 103 Italian provinces for the year 2003.

The spatial interaction between economic phenomena introduces the concept of spatial autocorrelation, which is linked to the territorial shape of the observed phenomena and to the connections between observations. Measures of spatial autocorrelation take into account the dependence between observations by a spatial weights matrix W. For a set of N observations the spatial matrix W is an $N \times N$ matrix with the diagonal elements equal to 0; the other elements w_{ij} represent the intensity of the effect of territorial area i on territorial area j (see Anselin and Bera 1998). The matrix defines the structure and the intensity of spatial effects, and it may be either a contiguity matrix or a matrix based on a distance decay function. In the literature, there are very few formal guidelines and suggestions on the choice of the most adequate spatial weights (for details, see Anselin 1988, 2002; Anselin and Bera 1998; Leenders 2002; Dietz 2002). Here, we use a rook contiguity matrix row-standardized, i.e., a binary spatial weight such that $w_{ij}^{s} = w_{ij}/\sum w_{ij}$ if the provinces i and j are contiguous (i.e., share a border), and $w_{ij} = 0$ otherwise. Though, other matrices could be used, in our view the contiguity matrix is the most appropriate to describe the spatial interactions of labour market in Italy, and to catch neighbouring effect in the local labour markets. Moreover, as the statistical units are territorial areas and not single points (e.g., families, firms, etc.), for example a generic distance matrix is less useful (see, Anselin 1988).[1]

[1] With regard to the provinces of the two islands Sardinia and Sicily, the contiguity has been considered inside each island.

In order to explore the significance of spatial clusters of high or low unemployment, our starting point is a cross-sectional regression model on regional unemployment without spatial effects:

$$U = \beta_0 + \beta_1 E + \sum_{k=2}^{K} \beta_k C_k + \varepsilon \tag{1}$$

where U is the log of the provincial unemployment rate; E is the log of provincial employment over working age population as a proxy for labour demand;[2] C_k are control variables; and ε is a vector of residuals. The control variables are: log of employment in the service sector over total provincial employment (E_{serv}); log of employment in the manufacturing sector over total provincial employment (E_{manif}); log of the size of the younger population (age from 15 to 29 years) over the total population (P_{15-29}) and log of the number of occupied houses over the total number of available houses (H_{occ}).

The variables E_{serv} and E_{manif} are proxies for the industry-mix explanation, though it is not always clear which sign these control variables should have; intuitively, provinces specializing in a declining economic sector such as manufacturing might show higher structural unemployment rates than provinces specializing in modern sectors such as services.

The size of the younger population with respect to the total population (P_{15-29}) is a proxy for natural change. Many studies have investigated whether the age structure of the population affects on the local unemployment rate. In the main, these studies have shown that areas with a relatively young population have a more stubborn unemployment problem, and that areas with a relatively old population experience a less persistent problem (Elhorst 1995; Molho 1995).

The H_{occ} is a proxy for economic and social barriers. The housing market where there is a lower proportion of occupied housing should have cheaper housing prices and more chance of finding housing compared with provinces where there is a high proportion of occupied housing. We will expect a negative sign of the variable, the reason being that workers are not available to move from area i with a high number of vacancies to province j with a low number of vacancies.

If the spatial effects are substantive but they are ignored, the OLS regression of Eq. (1) will provide a biased estimation of the parameters in the case of spatial lag dependence, while it provides unbiased and inefficient estimates in the case of spatial error dependence. Therefore, in order to explore the spatial interaction of the geographical distribution of unemployment, we follow loosely the robust specification strategy (see Anselin et al. 1996). Generally, one refers to the widest model including all kinds of spatial effects:

$$U = \rho \mathbf{WU} + \beta \mathbf{X} - \delta \mathbf{WX} + (\mathbf{I} - \lambda \mathbf{W})^{-1} \xi \tag{2}$$

[2] We use one-year lag of employment to avoid some possible endogeneity problems connected to this latter and the other explanatory variables.

where $\mathbf{X}$ is an $(n \times k)$ matrix of observations on the k independent variables (in our application the E and C_k variables). ρ is the spatial autocorrelation coefficient and measures the spillover effects: in other words, $\rho \neq 0$ implies that unemployment in province i is correlated to the unemployment in other neighbouring provinces. Moreover, in order to capture spillover effects connected to the explanatory variables, we include their spatial lags (i.e. C_k and E) with the coefficient δ.

Model 2 cannot be estimated directly because the parameters cannot be fully identified. So, in brief, according to the aforementioned strategy, first it is necessary to test whether it is appropriate to include autoregressive disturbances (Is λ =0?) or a spatially lagged dependent variable (Is ρ = 0?). Therefore, departing from a model without spatial effects (Model 1), using a separate robust Lagrange multiplier test (LM) (see Anselin et al. 1996) we test whether λ and ρ are equal to 0; if neither are equal to 0, we could choose between a spatial error or a spatial lag model, on the basis of which a robust LM statistic is larger. Moreover, it is worth asking whether a more general model would be preferable. In this case, an LR test on a common factor hypothesis should be done.[3]

In addition to this procedure, we follow a general empirical strategy according to Hendry's methodology (see, e.g., Spanos 1988). We distinguish between the theoretical model (i.e., the mathematical formulation of the theory, in our application Model 1) and the statistical models written in terms of observable random variables. If the assumptions of the statistical model are tested and not rejected, this indicates that the postulated probabilistic structure is appropriate for the data. If not, an alternative model, which has a more appropriate informative structure, must be chosen. In other words, we will try to maximize the *statistical adequacy* of the theoretical model. The empirical findings, discussed in the next section, were obtained in the light of this empirical strategy.

3 Empirical findings for provincial Italian unemployment

3.1 Preliminary findings

Although the previous section identified, on the basis of theory and the availability of the data, some relevant variables that explain the regional disparities of unemployment rates in Italy, it is not expected that all these variables (i.e the variables included in the estimable model) would be required in an adequate statistical model.

We first estimate a cross-sectional model without spatial effects. The estimations obtained are shown in column 1 of Table 1. All the coefficients of independent

[3] The autocorrelated error model is equivalent to a special form of spatial lag model by the following transformation of dependent and independent variables: $(\mathbf{Y} - \lambda\mathbf{WY})$ and $(\mathbf{X} - \lambda\mathbf{WX})$; so the spatial lag model can be written as $\mathbf{Y} = \lambda\mathbf{WY} + \mathbf{X}\beta - \lambda\mathbf{WX}\beta + \epsilon$. This is a subset, known as the common factor hypothesis model, of the more general model $\mathbf{Y} = \lambda\mathbf{WY} + \mathbf{X}\beta + \mathbf{WX}\delta + \epsilon$. The LR test of the common factor hypothesis tests the hypothesis $\delta = \lambda\beta$: if the null hypothesis is rejected a more general model with lagged independent variables must be estimated. Different procedures to test the significance of δ have been proposed (see Florax and Folmer 1992; Florax et al. 2003).

Table 1 Regression results

Variable	Col 1 Model 1 OLS	Col 2 Model 2 Lagged Dep. Var. MLE	Col 3 Model 3 OLS	Col 4 Model 4 Lagged Dep. Var. MLE	Col 5 Model 5 Autocorrelated error MLE	Col6 Model 6 Lagged Indip. Var Unlagged terms MLE	 Lagged terms MLE
E	−3.304 (0.00)	−2.496 (0.00)	−3.277 (0.00)	−2.399 (0.00)	−3.275 (0.00)	−2.406 (0.00)	−0.760 (0.57)
E_{serv}	0.037 (0.90)	0.457 (0.11)	–	–	–	–	–
E_{manif}	−0.446 (0.01)	−0.117 (0.52)	−0.472 (0.00)	−0.355 (0.00)	−0.442 (0.00)	−0.274 (0.03)	−0.093 (0.24)
P_{15-29}	0.359 (0.02)	0.080 (0.61)	0.394 (0.00)	0.309 (0.00)	0.372 (0.00)	0.116 (0.50)	−0.004 (0.97)
H_{occ}	0.014 (0.82)	0.007 (0.90)	–	–	–	–	–
λ	–	–	–	–	0.139 (0.29)	–	–
ρ	–	0.331 (0.00)	–	0.279 (0.00)	–	–	0.162 (0.19)
Log likelihood	−8.94	−2.15	−8.97	−3.42	−8.55	−1.54	–
LM(λ)	2.66 (0.10)	2.14 (0.14)	1.37 (0.24)	1.04 (0.31)	–	0.18 (0.67)	–
LM(ρ)	13.13 (0.00)	–	9.94 (0.00)	–	10.19 (0.00)	–	–
LR Common Factor	–	–	–	–	13.51 (0.01)	–	–
AIC	27.87	16.30	23.95	14.85	23.09	17.08	–

variables—except E_{serv} and H_{occ}—are statistically significant.[4] The reason why the coefficients of E_{serv} and H_{occ} are not statistically significant could be connected to the spatial correlation between observations, as is highlighted by LM test. The LM spatial test gives a significant value of 13.13 and this indicates that $\rho \neq 0$; so that a spatial lag model has to be estimated.

The estimations of a spatial lag regression model are shown in column 2 of Table 1. The coefficients of the variables—except the variable E are not statistically significant, but the coefficient of the variable WU is statistically significant and equal to $\rho = 0.33$. The positive value of ρ implies that unemployment in province i is correlated to the unemployment in other neighbouring provinces. In other words, provinces with high (or low) unemployment rates are clustered together. Moreover, the significant value of the coefficient of the variable E implies that unemployment in one area is strictly affected by the employment in the same area. Finally, the value of an LM error test of 2.14 with a p-value of 0.14 ($\lambda = 0$) implies that the residuals from the unemployment regression are not spatially autocorrelated.

The comparison between OLS and the spatial lag model (i.e. columns 1 and 2 of Table 1) highlights that the coefficients of the variables are very different. According to the diagnostic of spatial correlation, we should stop the analysis at the second step with Model 2 (column 2 of Table 1). This model eliminates the spatial correlation, but because nearly all the coefficients are not statistically significant, this leads us to investigate the spatial structure of each variable.

Could this lack of statistical significance of the coefficients be connected to the different spatial structure of the variables? In order to answer this question, we performed a more detailed analysis using exploratory spatial data analysis (ESDA) and other models. These estimates are reported in columns 3, 4, 5 and 6. The results are discussed in the next subsection.

3.2 Final results

In order to detect patterns of spatial association, spatial outliers or forms of spatial heterogeneity, we use some of the tools of ESDA. ESDA is a set of techniques aimed at: describing and visualizing spatial distribution; identifying atypical localizations or spatial outliers; detecting patterns or spatial association, clusters or hot spots; and suggesting spatial regimes or other forms of spatial heterogeneity (for details, see Haining 1990; Anselin 1998a,b). To know the spatial structure of variables and whether or not it is similar among variables will enable us—as with the time-structure in time series analyses—to identify the correct variables to include in a regression model. It could be that the lack of statistical significance of the coefficients in Model 2 (column 2 Table 1) might be connected with the different spatial structure of the variables considered.

According to Spanos (1988, p 117), the problem "*arises as to how to coalesce the relevant theoretical and sample information in the specification of statistical models*".

[4] It is noteworthy the not significant coefficient of H_{occ} could be connected to the rigidity on supply-side of Italian housing market. We would like to thank an anonymous referee for this suggestion.

Table 2 Moran's *I* statistics

Variable	Moran's I	Standardized value
U	0.856	12.595
E	0.863	12.691
E_{serv}	0.300	4.497
E_{manif}	0.530	7.848
P_{15-29}	0.852	12.533
H_{occ}	0.402	5.985

In other words, we need to identify an estimable model—with a theoretical basis—that is bound up with an adequate statistical model.

Table 2 displays the Moran's I statistic of all the variables. Inference is based on a standardized z-value that follows a normal distribution.

All the variables are positively spatially autocorrelated since the statistics are significant, with $p = 0.0001$ for every variable. Because, Moran's I is similar (but not equal) to a correlation coefficient, we could say that the variables show a different intensity of spatial association. This is higher for the variables U, E, E_{manif} and P_{15-29} than for the variables E_{serv} and H_{occ}. The high intensity of the global association index (Moran's I) indicates a tendency towards geographical clustering of similar provinces with a high (or low) value of the variable (e.g. provinces with high or low value of unemployment are geographically clustered). Conversely, the low positive value of Moran's I with regard to the variables E_{serv} and H_{occ} could indicate a non-geographical clustering of similar provinces; i.e. the low value of Moran's I indicates lack of similarity among provinces with respect to E_{serv} and H_{occ}.

The Moran's I statistic is a global statistic and does not allow us to investigate the provincial structure of spatial autocorrelation of each variable. It does not enable us to discover aspects such as: which provinces contribute more to the global spatial autocorrelation? Are there local spatial clusters of high or low values? If so, do these clusters identify a dual structure (North–South)? Do variables have the same or similar spatial heterogeneity?[5]

Therefore, a closer investigation of the spatial distribution of variables which explain the disparities of unemployment could be useful to identify the correct variables we need to include in the regression model. Maybe, in order to obtain an adequate statistical model, we have to include in the model variables with the similar spatial association and spatial structure. In order to explore the spatial distribution of our variables, the Moran scatterplot was used (see Anselin 1995).

The Moran scatterplot allows us to study the local spatial instability by plotting the spatially-lagged variable (e.g. WY) against the unlagged variable (e.g. Y). It may be subdivided into four quadrants corresponding to four spatial associations. The first, of

[5] As known the spatial effects are distinguished in spatial dependence and spatial heterogeneity. Spatial heterogeneity "may show up in terms of spatial heteroskedasticity or spatially varying parameters" (de Graaff et al. 2001, p. 259). As our data are not affected from heteroskedasticity, by using some ESDA tools we investigate the spatial structure. By a spatial regime model we investigated the possibility of spatial varying parameters; this last hypothesis has been rejected.

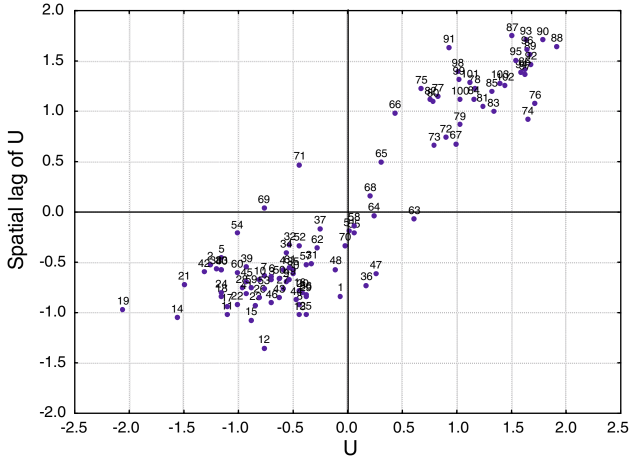

Fig. 1 Moran scatterplot of U

these associations on the top-right, are provinces with a large Y surrounded by large WY (quadrant HH). The second, on the top-left, are provinces with a small Y surrounded by large WY (quadrant LH). The third, on the bottom-left, are provinces with a small Y surrounded by small WY(quadrant LL). And the fourth, on the bottom-right, are provinces with a large Y surrounded by small WY (quadrant HL).[6]

The first and third quadrants (HH and LL) contain provinces with a positive spatial association, i.e. they indicate clusters of provinces with similar values. The second and fourth quadrants, however, display clusters of provinces with dissimilar values or a negative spatial association.

In Figs. 1, 2, 3, 4, 5 and 6, the Moran scatterplot for the variables U, E, E_{serv}, E_{manif}, P_{15-29} and H_{occ} is displayed. It can be seen (Fig. 1) that, with respect to the variable U, positive spatial association characterizes most Italian provinces: 91.3% of Italian provinces have a positive association or similar values (35.0% in quadrant HH, and 56.3% in quadrant LL). With respect to the variable E, it has an equal but opposite spatial structure in comparison with the unemployment variable. The percentage of provinces with a positive association is again equal to 91.3% (56.3% in quadrant HH, and 35.0% in quadrant LL: see Fig. 2).

The same could be done with respect to the variables E_{manif} and P_{15-29} (Figs. 4 and 5); they present a spatial structure very similar to that of U and E. In particular, with respect to E_{manif}, 80.1% of provinces present a positive spatial association, whereas with respect to P_{15-29}, 91.3% have a positive spatial association. In contrast to the previous variables, the variables E_{serv} and H_{occ} do not show a high positive spatial association. In fact the percentage of provinces with positive association is equal to 65.0% for E_{serv} and 69.9% for H_{occ} (Figs. 3 and 6).

Moreover, the Moran scatterplots help us to identify atypical provinces, i.e. those provinces characterized by an association of dissimilar values. In particular, the Moran

6 A value is large or small with respect to its average value.

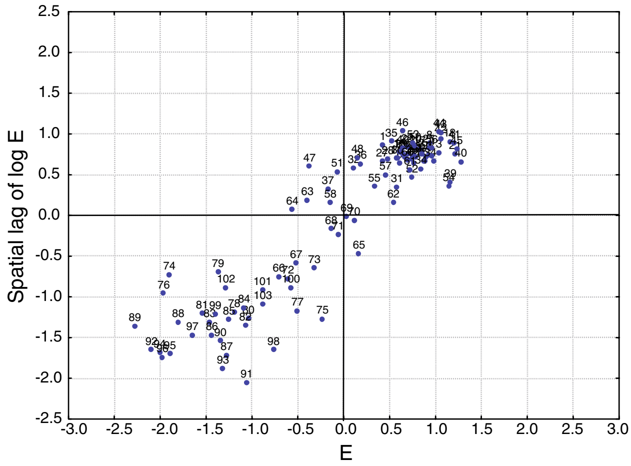

Fig. 2 Moran scatterplot of E

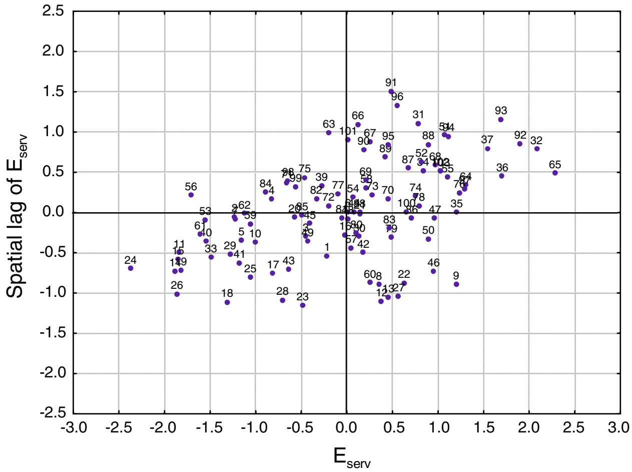

Fig. 3 Moran scatterplot of E_{serv}

scatterplot of the dependent variable U shows 2 and 7 provinces in the second (LH) and fourth (HL) quadrant, respectively (i.e. 9 in total). Similarly, also in the Moran scatterplot of variables E, E_{manif} and P_{15-29}, there are, respectively, 9, 9 and 20 provinces in the HL and LH quadrants taken together. Conversely, there are more provinces deviating from the global pattern of positive correlation for the variables E_{serv} and H_{occ} than for the others, viz. there are 26 provinces with an association of dissimilar values in the case of E_{serv}, and 31 provinces in the case of H_{occ}.

All these results lead us to the conclusion that there is a different spatial structure among variables. This is similar or almost identical for the variables U, E, E_{manif} and P_{15-29}, but not for the variables E_{serv} and H_{occ}. The Moran scatterplot of variables such as E, E_{manif} and P_{15-29} indicates the presence of spatial heterogeneity that may

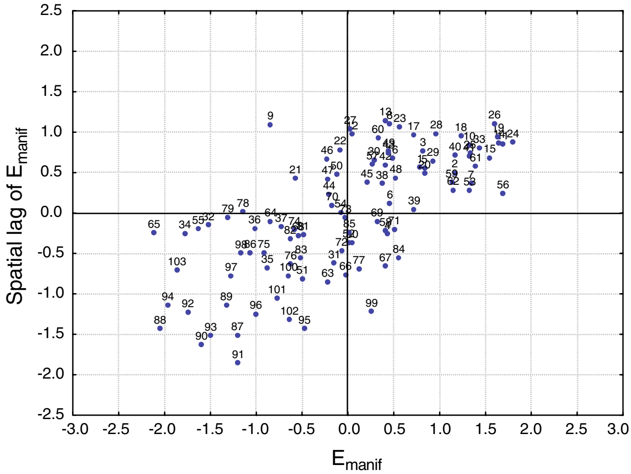

Fig. 4 Moran scatterplot of E_{manif}

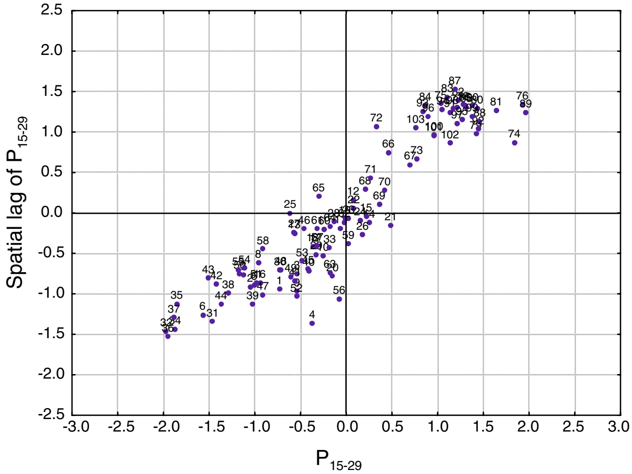

Fig. 5 Moran scatterplot of P_{15-29}

be synthesized in four distinct clusters. The first corresponds to the HH scheme and includes mainly the Central-northern provinces. The second corresponds to the LL scheme and includes mainly the Southern provinces. Both regimes have a positive spatial association. The third and the fourth regimes correspond, respectively, to the LH and the HL scheme which both exhibit an atypical negative spatial association and include mainly Central and Southern provinces.

The Moran scatterplot of variable U indicates the presence of spatial heterogeneity that may be subdivided into three clusters. The first of these corresponds to the HH scheme, including mainly Southern provinces. The second corresponds to the LL scheme and includes mostly Central-northern provinces. Both regimes exhibit a

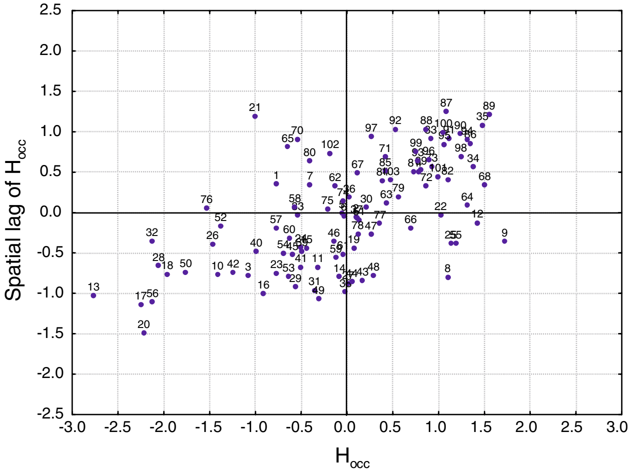

Fig. 6 Moran scatterplot of H_{occ}

positive spatial association. The third corresponds to the HL scheme which has only seven provinces with an atypical negative spatial association.

The exploratory spatial data analysis of the distribution of the variables U, E, E_{manif} and P_{15-29} highlights a Central-northern and Southern spatial polarization scheme; moreover these variables are characterized by a high value of I-Moran statistic.

In contrast to the above, the other independent variables, E_{serv} and H_{occ} show a lower I-Moran value and a spatial structure that cannot be encapsulated in specific spatial pattern. Specifically, with respect to spatial structure aspect, the quadrants LL and HH include both Central-northern and Southern provinces together; i.e. there is not a sharp pattern as the above variables. Further, the large number of provinces contained in the LH and the HL scheme is an expression of *pockets of nonstationarity*.

In the light of the above implications, the different spatial structure between both the dependent variable and the independent variables E_{serv} and H_{occ}, and between the latter two variables and the other independent variables could explain why most of the coefficients of Model 2 of Table 1 are not statistical significant. Therefore, we have estimated a new OLS model leaving out the variables E_{serv} and H_{occ}.

Column 3 of Table 1 shows the results of a cross-sectional model that neglects spatial effects. The coefficients have the expected sign and they are statistically significant. The statistically significant value of 9.94 of the LM spatial lag test indicates a value of $\rho \neq 0$, so a spatial lag model has to be estimated.

In column 4 of Table 1 the estimations of the spatial lag model are shown. All the coefficients are statistically significant and have the expected sign. The value of the LM error test of 1.04 with a p-value of 0.31 ($\lambda = 0$) implies that the residuals from the unemployment regression are not spatially autocorrelated.

The estimations highlight that strong neighbouring effects occurred in 2003 among Italian provinces, as shown by the high and significant value of ρ (about 0.28). Further, the unemployment differences are also explained by the demand-side variables; specifically, the coefficient of E is high and significant as well as the coefficient of

E_{manif}. A marginal increase in employment produces a more proportional decrease in unemployment, while a marginal increase in employment in the manufacturing sector is translated into a less proportional decrease in unemployment. With respect to the last (E_{manif}), we note here that, notwithstanding manufacturing is a decline sector, it has a reducing effect on unemployment rate.

On the supply-side, the demographic variable (P_{15-29}) also has the expected sign but, contrary to expectations, a marginal increase of younger people has not very strong effect on unemployment, i.e. unemployment increases less proportionally.

Although, the LM error test does not indicate the presence of spatial correlation in the residuals, before choosing the best statistical model, we estimate an autocorrelated error model.

The estimations of the autocorrelated error model show that all coefficients are statistically significant and with the expected signs. The LM spatial lag test with the significant value of 10.19 indicates a value of $\rho \neq 0$ that points us to a spatially-lagged dependent variable model. Moreover, the LR test of common factor hypothesis rejects the null hypothesis, so a model with lagged independent variables must be estimated.

Column 6 of Table 1 shows the estimation of the model with lagged independent variables: on the left of column 6 the coefficients of the independent variables are shown, and on the right the coefficients of the lagged independent variables. The diagnostic of spatial dependence shows that no spatial autocorrelation remains in the errors.

Among the estimated models, we prefer the spatial lag model (column 4 of Table 1) because it is most parsimonious and best puts together the relevant theoretical insight with the sample information. It has one of the smallest values of AIC (Akaike Information Criteria) and is the best model according to the spatial diagnostic. Moreover, we still preferred the spatial lag model over the spatial error model, notwithstanding the last shows also significant coefficients and no spatial correlation in the residuals. In fact, the evaluation of the spatial dependence of the latter model is the expression of the joint effect of omitted variables, the model misspecification and spatial autocorrelation that cannot be disentangled; while in the spatial lag model spatial dependence is measured by the coefficient ρ which deserves an economic interpretation.

4 Conclusions

In this paper, using some tools ESDA and spatial econometrics, the most adequate statistical model has been found in order to explain the provincial unemployment differences of the labour market in Italy for the year 2003. Specifically, by ESDA we investigated on the probabilistic structure of spatial data identifying those variables with a similar spatial structure to be inserted into the statistical model.

There are two conclusions worthy to attention. The first is related to the ESDA, if it is applied to all variables, dependent and independent, it could be fruitfully used like an integration test on time series. In other words, it is let to have more information on the spatial structure of variables before insert them in the model.

The second insight is related to the economic implications connected on the neighbouring effect coming out from the spatial lag model. The estimates show that the

Italian labour market, in 2003, is characterized by a polarization of unemployment. In other words, a neighbouring effect features the Italian provincial unemployment, that is local labour markets with high or low values of unemployment tend to cluster in space (i.e. positive spatial autocorrelation).

These differences are strongly explained by the labour demand variables. This is in accordance to previous analysis on Italian labour market performed by different methods and for different years (see, e.g., Amendola et al. 1999; Amendola et al. 2004).

The strong effect of demand variable on the unemployment differences is justified by the theory as well as the similar spatial structure of two variables (i.e. U and E). In fact, ESDA shows that labour demand is featured alike the unemployment by a strong polarization. This result is not surprising rather it is consisting to other results on European regions (Overman and Puga 2002). In accordance to Puga (2002), we believe the polarization of employment reflects the clustering of activities, which is now present despite a high share of European Structural Fund Expenditure has been addressed to Objective 1 regions to even out the growth differences and consequently unemployment and employment differences among territorial areas (see, e.g. Cracolici et al. 2007).

Moreover, the unemployment differences are, to a much lesser extent, explained by the share of young population and economic structure composition. The positive sign of coefficient of the former variable highlights an inefficiency and rigidity of Italian local labour markets to assimilate the younger cohorts that likely, on average, have higher level of education and skills (see Rodríguez-Pose and Fratesi 2004). The significant and negative coefficient of employment in manufacturing sector points to decreasing regional unemployment rates notwithstanding its share on total economy is declining.

In brief, the empirical results let state recent policies on labour market carry a reduction of the nation-wide unemployment rate leaving a high unemployment differentiation across provinces; indeed, the low flexibility of the labour market has allowed some northern provinces to achieve higher employment, while the provinces in the southern have only modestly decreased their unemployment rates.

In conclusion, the polarized spatial pattern of regional unemployment rates implies that the effect on one province spread over to the neighbouring ones, so policy makers have to pay attention on it because a regional intervention will benefit neighbouring regions.

References

Amendola A, Caroleo FE, Coppola G (1999) Differenziali territoriali nel mercato del lavoro e sviluppo in Italia. In: Biagioli M, Caroleo FE, Destefanis S (eds) Struttura della contrattazione. Differenziali salariali e occupazione in ambiti regionali. ESI, Napoli

Amendola A, Caroleo FE, Coppola G (2004) Divari regionali in Europa, Discussion Paper 78 CELPE Salerno
Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht
Anselin L (1992) SpaceStat tutorial. A workbook for using SpaceStat in the analysis of spatial data. University of Illinois, Urbana-Champaign
Anselin L (1992) SpaceStat tutorial. A workbook for using SpaceStat in the analysis of spatial data. University of Illinois, Urbana-Champaign
Anselin L (1995) Local indicators of spatial association-LISA. Geogr Anal 27:93–115
Anselin L (1998a) Interactive techniques and exploratory spatial data analysis. In: Longley PA, Goodchild MF, Maguire DJ, Wind DW (eds) Geographical information systems: principles, techniques, management and applications. Wiley, New York
Anselin L (1998b) Exploratory spatial data analysis in a geocomputational environment. In: Longley PA, Brooks SM, McDonnell R, Macmillan B (eds) Geocomputation, a primer. Wiley, New York
Anselin L (2002) Under the hood. Issues in the specification and in the interpretation of spatial regression model. Agric Econ 27:247–267
Anselin L, Bera AH, Florax R, Yoon MJ (1996) Simple diagnostic test for spatial dependence. Reg Sci Urban Econ 26:77–104
Anselin L, Bera A (1998) Spatial dependence in linear regression models with an application to spatial econometrics. In: Ullah A, Giles DEA (eds) Handbook of applied economic statistics. Springer, Berlin
Aragon Y, Haughton D, Haughton J, Leconte E, Malin E, Ruiz-Gazen A, Thomas-Agnan C (2003) Explaining the pattern of regional unemployment: the case of the Midi-Pyrénées region. Pap Reg Sci 82:155–74
Brunello G, Lupi C, Ordine P (2001) Widening differences in Italian regional employment. Labour Econ 8:103–129
Contini B, Trivellato U (2005) Eppur si muove. Dinamiche e persistenze nel mercato del lavoro Italiano. Il Mulino, Bologna
Cracolici MF, Cuffaro M, Nijkamp P (2007) Geographical distribution of unemployment: An analysis of provincial differences in Italy. Growth Change 38(4):642–665
Decressin J, Fatás A (1995) Regional labour dynamics in Europe. Euro Econ Rev 39:1627–1695
Graaff Tde , Florax RJGM, Nijkamp P, Reggiani A (2001) A general misspecification test for spatial regression models: dependence, heterogeneity and nonlinearity. J Reg Sci 41(2):255–276
Dietz RD (2002) The estimation of neighborhood effects in the social sciences: an interdisciplinary approach. Soc Sci Res 31:539–575
Dohse D, Krieger-Boden C, Soltwedel R (2002) EMU and regional labor market disparities. In: Cuadrado-Roura J, Parellada M (eds) Regional convergence in the European union. Springer, New York
Eichengreen B (1992) Labor markets and European monetary unification. Mimeo, University of California, Berkeley, CA
Elhorst JP (1995) Unemployment disparities between regions in the European union. In: Armstrong HW, Vickerman RW (eds) Convergence and divergence among European regions. Pion, London
Elhorst JP (2003) The mystery of regional unemployment differentials: theoretical and empirical explanations. J Econ Surv 17:709–748
Faini R, Galli G, Gennai P, Rossi F (1997) An empirical puzzle: falling migration and growing unemployment differential among Italian regions. Euro Econ Rev 41:571–579
Fischer M, Nijkamp P (1987) Regional labour markets. Elsevier, Amsterdam
Florax RJGM, Folmer H (1992) Specification and estimation of spatial linear regression models: Monte Carlo evaluation and pre-estimators. Reg Sci Urban Econ 22:405–432
Florax RJGM, Folmer H, Rey SJ (2003) specification searchers in spatial econometrics: the relevance of Hendry's methodology. Reg Sci Urban Econ 33:557–579
Gambarotto F, Maggioni MA (2002) Localizzazione e agglomerazione produttiva: L'analisi territoriale dei mercati del lavoro, In: Proceedings of the Italian association of labour economics vol XVII Scientific Meeting
Haining R (1990) Spatial data analysis in the social and environment sciences. Cambridge University Press, Cambridge
Hall R (1970) Why is the unemployment rate so high at full employment?. Brook Pap Econ Act 3:369–402
Jimeno JF, Bentolila S (1998) Regional unemployment persistence (Spain, 1976–1994). Labour Econ 5:25–51

Leenders RThAJ (2002) Modelling social influence through network autocorrelation: constructing the weight matrix. Soc Netw 24:21–47
López-Bazo E, del Barro T, Artis M (2002) The regional distribution of Spanish unemployment: a spatial analysis. Pap Reg Sci 81:365–389
López-Bazo E, del Barro T, Artis M (2005) Geographical distribution of unemployment in Spain. Reg Stud 39(3):305–318
Marston S (1985) Two views of the geographic distribution of unemployment. Q J Econ 100:57–69
Martin R (1997) Regional unemployment disparities and dynamics. Reg Stud 31:237–252
Molho I (1995) Spatial autocorrelation in British unemployment. J Reg Sci 35:641–658
Niebuhr A (2003) Spatial interaction and regional unemployment in Europe. Euro J Spat Dev 5:2–24
Overman HG, Puga D (2002) Unemployment cluster across Europe's regions and countries. Econ Policy 34:116–147
Patridge MD, Rickman DS (1997) The dispersion of US State unemployment rates: the role of market and non-market equilibrium factors. Reg Stud 31(6):503–606
Prasad ES, Utili F (1998) The Italian labor market: stylized facts, institutions, and directions for reform. IMF Working Paper
Puga D (2002) European regional policies in light of recent location theories. J Econ Geog 2:373–406
Rodríguez-Pose A, Fratesi U (2004) Between development and social policies: the impact of European structural founds in Objective 1 regions. Reg Stud 38:97–113
Spanos A (1988) Towards a unifying methodological framework for econometric modelling. Econ Notes 1:107–141